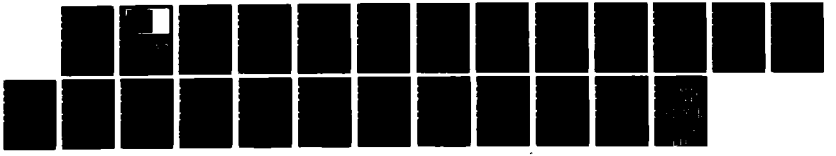
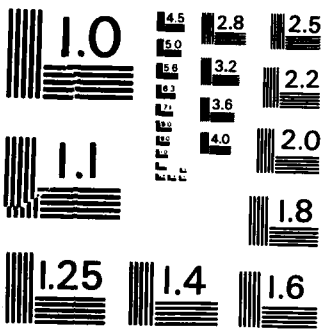
AD-A163 628 MAXIMUM LIKELIHOOD RECURSION AND STOCHASTIC APPROXIMATION IN SEQUENTIAL DESIGNS(U) WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER C F WU NOV 85 1/1

UNCLASSIFIED MRC-TSR-2890 DAAG29-80-C-0041 F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

(2)

MRC Technical Summary Report #2890

MAXIMUM LIKELIHOOD RECURSION AND STOCHASTIC APPROXIMATION IN SEQUENTIAL DESIGNS

C. F. J. Wu

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

November 1985

(Received July 10, 1985)

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

MAXIMUM LIKELIHOOD RECURSION AND STOCHASTIC APPROXIMATION IN SEQUENTIAL DESIGNS*

C. F. J. Wu

Technical Summary Report #2890
November 1985

ABSTRACT

An approach to sequential design for estimating the root of a nonlinear equation is described. It sets the next design point at the current estimate of the parameter via a parametric model and maximum likelihood (or other efficient) estimation. For normal, binomial and Poisson errors and their respective canonical link functions, it is close to the Robbins-Monro stochastic approximation and thus enjoys the latter's robustness against the misspecification of the link function. Some new variations of the Robbins-Monro scheme are obtained as a consequence.

AMS (MOS) Subject Classifications: 62L20, 62L05

Key Words: maximum likelihood recursion, stochastic approximation, Robbins-Monro, binary experiment, Poisson experiment, generalized linear model, canonical link

Work Unit Number 4 (Statistics and Probability)

---

*To appear in "Adaptive Statistical Procedures and Related Topics" (J. Van Ryzin, ed.), 1986.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the A. P. Sloan Foundation.

SIGNIFICANCE AND EXPLANATION

When the saving of sample size is an important consideration, sequential design of experiments is often used. By efficiently utilizing the information in the past experiments, it determines how the next experiment should be conducted. Statistical theory for sequential designs has been developed for normal and binomial variations. For the problem of determining the solution of an unknown nonlinear equation, we have developed a class of sequential design procedures that can handle very general variations described by the generalized linear models. In special cases it includes a new adaptive version of the Robbins-Monro stochastic approximation and a maximum likelihood recursion scheme for quantal responses. Its relation to the stochastic approximation and the role the link function plays are studied. Theoretical issues such as consistency, robustness, asymptotic normality and second-order properties are discussed.

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

MAXIMUM LIKELIHOOD RECURSION AND STOCHASTIC APPROXIMATION IN SEQUENTIAL DESIGNS*

C. F. J. Wu

1. INTRODUCTION

We are interested in efficient sequential designs for estimating the root of an unknown nonlinear equation, where the distribution of the responses is quite general (continuous or discrete). The proposed approach is based on design updating with the maximum likelihood estimate via parametric model. It is dubbed the maximum likelihood (ML) recursion approach. In several important situations it is shown to be closely related to the stochastic approximation approach of Robbins and Monro (1951).

The problem can be described as follows. The response y is related to an underlying "design" variable x. Denote the mean of y at x by $M(x)$, which is unknown to the experimenter. Of special interest is the solution x^* to the equation $M(x) = p$. In bioassay x^* may be an effective dose level; in control system x^* may be an optimal input level. Usually the distributional form of y is roughly known, e.g., binomial in bioassay. Denote it by $f(y|\theta)$. The parameter θ is related to x via a link function unknown to the experimenter. For univariate x, our approach starts with assuming a parametric link function $\theta = g(\lambda x - \alpha)$ with g known. An efficient estimate $(\hat{\lambda}_n, \hat{\alpha}_n)$ of (λ, α) is obtained based on the first n observations via the assumed model f and g. Under f and g, $E(y|x)$ is

---

*To appear in "Adaptive Statistical Procedures and Related Topics" (J. Van Ryzin, ed.), 1986.

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041 and the A. P. Sloan Foundation.

a function of α and λ, denoted by $H(x|\alpha,\lambda)$. It is typically monotone in x. The ML recursion chooses the next design x_{n+1} to satisfy $H(x_{n+1}|\hat{\alpha}_n,\hat{\lambda}_n) = p$. The procedure can be repeated indefinitely. The idea was first studied in Wu (1985) for binary y.

In Section 2 this approach for normal error and linear link function is shown to yield a nonadaptive Robbins-Monro (RM) stochastic approximation (1) if λ in the preceding description is fixed. For α and λ unknown, it leads to a variation (11) of an adaptive RM scheme. The new scheme (11) has the same first order asymptotic behavior as the adaptive RM. But the second order behavior, yet to be investigated, will probably be different. The RM scheme, without any knowledge of M, has desirable asymptotic properties under weak conditions on M. This robustness is therefore shared by the ML recursion, although the latter is based on the assumption of a possibly incorrect link function $g(\lambda x - \alpha)$. Its robustness, in the case of binary data with logit link, is shown to stem from the iteration step $H(x_{n+1}|\hat{\alpha}_n,\hat{\lambda}_n) = p$. See (16) and (17). Section 3 contains other results that link the two approaches. A general description of the ML recursion approach for generalized linear models is given in Section 4. Canonical link functions are recommended. In the case of Poisson variation, the ML recursion based on a canonical link is equivalent to a version of the RM recursion (1). This connection enables us to study the asymptotic behavior of the ML recursion. The paper concludes with the pros and cons of the ML recursion approach relative to the RM recursion approach and points out potential gains in relating the two approaches. No rigor is attempted throughout the paper.

2. ROBBINS-MONRO PROCEDURE AND ITS VARIATIONS VIA LEAST SQUARES RECURSION

In their pioneering paper Robbins and Monro (1951) proposed the following recursive scheme

$$x_{n+1} = x_n - \frac{c}{n}(y_n - p), \tag{1}$$

for estimating the solution x^* of $M(x^*) = p$, where the observation y_n taken at x_n satisfies $y = M(x) + \varepsilon$ with $E(\varepsilon) = 0$. The scheme does not assume any knowledge of M, which is typically unknown. Under weak conditions on M and ε, x_n is known to converge to x^* with probability one as $n \to \infty$ (Robbins and Siegmund, 1971) and to be asymptotically normal (Sacks, 1958). The optimal choice of c for minimizing the asymptotic variance of x_n is $(M'(x^*))^{-1}$, $M'(x^*) \neq 0$.

The Robbins-Monro (RM) scheme (1) can be interpreted as a recursive scheme with least squares updating. Let us make a tentative assumption that $y = \alpha + \beta x + \varepsilon$ and the errors ε have mean zero, variance σ^2 and are uncorrelated. First we consider the simple case of known β. For solving the linear equation $\alpha + \beta x = 0$ (p is now zero), the parameter of interest is $\theta = -\alpha/\beta$. Based on the first n observations, the least squares estimate (or the maximum likelihood estimate if the errors are normal) of θ is

$$\hat{\theta}_n = -\frac{\hat{\alpha}_n}{\beta} = \bar{x}_n - \frac{\bar{y}_n}{\beta}, \tag{2}$$

where $\bar{x}_n$ and $\bar{y}_n$ are respectively the means of x_i and y_i, $i = 1(1)n$. If the next observation y_{n+1} is taken at the current estimate $\hat{\theta}_n$ of θ, the recursive relation

$$x_{n+1} = \hat{\theta}_n = \bar{x}_n - \frac{\bar{y}_n}{\beta} \tag{3}$$

obtains. It is easy to see that (3) is equivalent to

$$x_{n+1} = x_n - \frac{\beta^{-1}}{n}(y_n - 0), \tag{4}$$

which is the RM recursion (1) with $c = \beta^{-1}$. This equivalence was pointed out by Lai and Robbins (1979). It is a significant step since it connects two seemingly distinct approaches to the design problem outlined in Section 1. The approach that leads to (3) is parametric in that it is motivated by a linear function that links $E(y)$ and x and, to a lesser extent, by the normality of errors (which makes the least squares estimator fully efficient). On the other hand, the stochastic approximation approach (4) is nonparametric in that its asymptotic performance is very much independent of the knowledge of $M(x)$. The assumption $y = \alpha + \beta x + \varepsilon$ is useful for motivating and generating design procedures. The validity and performance of the resulting design are nonetheless independent of the assumption.

So far we have assumed that the slope parameter β is known. For unknown β, what recursive scheme will the least squares updating approach lead to? Here

$$x_{n+1} = \hat{\theta}_n = -\frac{\hat{\alpha}_n}{\hat{\beta}_n} = \bar{x}_n - \frac{\bar{y}_n}{\hat{\beta}_n}, \tag{5}$$

and

$$\hat{\beta}_n = \sum_1^n y_i(x_i - \bar{x}_n) \Big/ \sum_1^n (x_i - \bar{x}_n)^2 \tag{6}$$

is the regression slope estimate. From (5) and (7)

$$\bar{x}_n = \bar{x}_{n-1} + (x_n - \bar{x}_n)/(n-1) = \bar{x}_{n-1} + (x_n - \bar{x}_{n-1})/n, \tag{7}$$

we have the following recursive relation

$$\begin{aligned} x_{n+1} - x_n &= \bar{x}_n - \bar{x}_{n-1} - (\bar{y}_n/\hat{\beta}_n - \bar{y}_{n-1}/\hat{\beta}_{n-1}) \\ &= (x_n - \bar{x}_{n-1})/n + \bar{y}_{n-1}/\hat{\beta}_{n-1} - \bar{y}_n/\hat{\beta}_n \\ &= \left(1 - \frac{1}{n}\right)\bar{y}_{n-1}/\hat{\beta}_{n-1} - \bar{y}_n/\hat{\beta}_n . \end{aligned} \tag{8}$$

The last equality follows from (5) with n replaced by $n - 1$. By using (7)

for y, this equals

$$(9) \qquad (1 - \tfrac{1}{n})\bar{y}_{n-1}(\hat{\beta}_{n-1}^{-1} - \hat{\beta}_n^{-1}) - y_n/(n\hat{\beta}_n) \ .$$

It remains to derive an explicit form for the first term of (9). From (7.2.7) of Goodwin and Payne (1977),

$$\hat{\beta}_n - \hat{\beta}_{n-1} = \frac{\Sigma_1^{n-1}(x_n - x_i)(y_n - \bar{y}_{n-1} - \hat{\beta}_{n-1}(x_n - \bar{x}_{n-1}))}{(n-1)\Sigma_1^{n-1}(x_i - \bar{x}_{n-1})^2 + \Sigma_1^{n-1}(x_i - x_n)^2}$$

$$= \frac{(n-1)(x_n - \bar{x}_{n-1})y_n}{n\Sigma_1^n(x_i - \bar{x}_n)^2} = d_{n-1} \ .$$

The second equality follows from (5). Therefore

$$(1 - \tfrac{1}{n})\bar{y}_{n-1}\left(\frac{1}{\hat{\beta}_{n-1}} - \frac{1}{\hat{\beta}_n}\right) = \frac{n-1}{n}\bar{y}_{n-1}\frac{d_{n-1}}{\hat{\beta}_{n-1}\hat{\beta}_n}$$

$$(10) \qquad = \frac{y_n}{n\hat{\beta}_n}\frac{\bar{y}_{n-1}}{\hat{\beta}_{n-1}}\frac{(n-1)^2(x_n - \bar{x}_{n-1})}{n\Sigma_1^n(x_i - \bar{x}_n)^2}$$

$$= -\frac{y_n}{n\hat{\beta}_n}\frac{(n-1)^2(\bar{y}_{n-1}/\hat{\beta}_{n-1})^2}{n\Sigma_1^n(x_i - \bar{x}_n)^2} \ .$$

The last equality follows from (5). From (8)-(10) follows a new recursive scheme

$$(11) \qquad x_{n+1} = x_n - \frac{\hat{\beta}_n^{-1}}{n}\left[1 + \frac{(n-1)^2(\bar{y}_{n-1}/\hat{\beta}_{n-1})^2}{n\Sigma_1^n(x_i - \bar{x}_n)^2}\right]y_n \ .$$

If the second term inside the square bracket is ignored, (11) reduces to the RM recursion (1) with $c = \hat{\beta}_n^{-1}$ and $p = 0$. Such an <u>adaptive</u> procedure with proper truncation on $\hat{\beta}_n$ was shown (Lai and Robbins, 1981; Anbar, 1978) to have the same minimal asymptotic variance as the optimal choice $c = (M'(x^*))^{-1}$.

We now study the order of magnitude of the "correction term"

$$(12) \qquad \frac{(n-1)^2(\bar{y}_{n-1}/\hat{\beta}_{n-1})^2}{n\Sigma_1^n(x_i - \bar{x}_n)^2},$$

which, being positive, makes the adjustment $|x_{n+1} - x_n|$ in (11) bigger than that in the adaptive RM scheme mentioned above. One can argue heuristically from the results of Lai and Robbins (1981) that $(n-1)(\bar{y}_{n-1})^2 = O_p(1)$, $\hat{\beta}_n \to M'(x^*)$, and $\sum_1^n(x_i - \bar{x}_n)^2 = O_p(\log n)$. The correction term (12) is therefore of the order $O_p((\log n)^{-1})$. We conjecture that the scheme (11) with proper truncation on the coefficient of $n^{-1}y_n$ has the same limiting distribution as the optimal RM scheme (1) with $c = (M'(x^*))^{-1}$ and the adaptive RM scheme (1) with $c = \hat{\beta}_n^{-1}$. If y_i is related to x_i by the simple linear regression model $y_i = \alpha + \beta x_i + \varepsilon_i$, this was established in Lai and Robbins (1982) with a different truncation scheme. What the correction term (12) does to (11) is in the lower order terms. Expand the mean square error of x_n as

$$(13) \qquad E(x_n - x^*)^2 = \frac{a_1}{n} + \frac{a_2}{nb_n} + \text{lower order terms},$$

where $b_n \to \infty$ as $n \to \infty$. The a_1 term is the same for the three procedures, while the a_2 term may differ. Since the scheme (11) is based on the least squares estimator, it may be second-order optimal (in an appropriate sense) for nearly linear $M(x)$ and normal errors. Second order asymptotic results, currently unavailable in the literature, may provide further insights into those small or moderate sample phenomena not readily explainable by first-order theory. Such results are found in section 6 of Wu (1985). Of course, small sample behavior depends on the location of initial observations.

The correction term (12) is non-negligible only for small or moderate n when $\log n$ is not large, $|\bar{y}_n| >> 0$, or $x_1,\ldots,x_n$ are not wide-spread. To make the scheme (11) more robust against poor choice of the starting value x_0 and the motivating linearity and normality assumptions on M and ε, the correction term (12) should be made less dependent on the remote past. Such can be achieved by replacing $\bar{y}_{n-1}$, $\hat{\beta}_{n-1}$ and $\Sigma(x_i - \bar{x}_n)^2$ in (12) by weighted versions with more weights on recent observations.

3. MAXIMUM LIKELIHOOD RECURSION IN BINARY EXPERIMENTS

In a binary experiment the outcome y is denoted by 1 (response) or 0 (nonresponse). The probability of response is related to a stress level x (at which the experimentation is performed) by

$$M(x) = \text{Prob}\{y = 1|x\} = E(y|x) .$$

It is often of interest to estimate the percentile x^* of $M(x)$, i.e. $M(x^*) = p$, $0 < p < 1$. The problem is that in practice the form of M is often unknown. For expensive runs, sequential experimentation, if feasible in practice, is called for since the data can be collected and used in a most effective way. For related comments, see Wu (1985).

The maximum likelihood (ML) recursion approach starts with a parametric model for the unknown $M(x)$. First we consider a one-parameter model $H(x - \alpha)$ with parameter α and H known. For estimating the $100p^{th}$ percentile, H is chosen to satisfy $H(0) = p$. That is, if H is the true model, α is the $100p^{th}$ percentile of M. The log likelihood for the first n observations is

$$\sum_1^n y_i \log H(x_i - \alpha) + \sum_1^n (1 - y_i)\log(1 - H(x_i - \alpha)) .$$

The maximum likelihood estimate $\hat{\alpha}_n$ of α satisfies the equation

$$\sum_1^n y_i \frac{H'(x_i - \alpha)}{H(x_i - \alpha)(1 - H(x_i - \alpha))} = \sum_1^n \frac{H'(x_i - \alpha)}{1 - H(x_i - \alpha)} .$$

By writing

$$K(x) = \frac{H'(x)}{H(x)(1 - H(x))} ,$$

the above equation can be expressed as a <u>weighted</u> normal equation

$$\sum_1^n y_i K(x_i - \hat{\alpha}_n) = \sum_1^n H(x_i - \hat{\alpha}_n)K(x_i - \hat{\alpha}_n) .$$

According to the ML recursion approach, the next design x_{n+1} is chosen to be the current estimate $\hat{\alpha}_n$ of α and the preceding equation becomes

$$(14) \qquad \sum_1^n y_i K(x_i - x_{n+1}) = \sum_1^n H(x_i - x_{n+1})K(x_i - x_{n+1}) \ .$$

To obtain a recursive relation between x_{n+1} and x_n, (14) gives

$$\sum_1^n [H(x_i - x_{n+1})K(x_i - x_{n+1}) - H(x_i - x_n)K(x_i - x_n)] + pK(0)$$

$$= \sum_1^n y_i [K(x_i - x_{n+1}) - K(x_i - x_n)] + y_n K(0)$$

$$\doteq \sum_1^n y_i K'(x_i - x_n)(x_n - x_{n+1}) + y_n K(0) \ .$$

Unless $K' = 0$, $x_{n+1} - x_n$ depends on all the past x_i and y_i. Recall that the RM recursion depends on the past $\{y_i\}$ through y_n. Only when $K' = 0$, $x_{n+1} - x_n$ behaves more like the RM scheme in this regard. Note that $K' = 0$ iff H is of the logistic form $\left(1 + \left(\frac{1}{p} - 1\right)e^{-cx}\right)^{-1}$. Another advantage of the logit-based ML recursion design is that it is less susceptible to poor choice of initial observations than the probit-based procedure (Sellke, 1986).

Without loss of generality, assume $c = 1$. For the logit assumption, equation (14) takes the form

$$\sum_1^n y_i = \sum_1^n \left(1 + \left(\frac{1}{p} - 1\right)e^{-(x_i - x_{n+1})}\right)^{-1} \ ,$$

which yields the recursive relation

$$y_n = \sum_1^n \frac{d\left(e^{-(x_i - x_n)} - e^{-(x_i - x_{n+1})}\right)}{\left(1 + de^{-(x_i - x_n)}\right)\left(1 + de^{-(x_i - x_{n+1})}\right)} + p, \quad d = \frac{1}{p} - 1 \ ,$$

or equivalently,

$$(15) \qquad \sum_{1}^{n} \frac{de^{x_n - x_i}(1 - e^{x_{n+1} - x_n})}{(1 + de^{x_n - x_i})(1 + de^{x_n - x_i} e^{x_{n+1} - x_n})} = y_n - p \ .$$

The special case $p = \frac{1}{2}$ was given in (13) of Wu (1985). The recursion (15) defines $x_{n+1} - x_n$ implicitly as a nonlinear function of $y_n - p$ and $\{x_i\}_1^n$. It is qualitatively similar to the RM scheme in two regards. It pushes x_n in the "right" direction, i.e., $x_{n+1} - x_n$ has the same sign as $p - y_n$. The step size $|x_{n+1} - x_n|$ gets smaller as n increases (and eventually at the rate n^{-1}).

The ML recursion approach can also be applied to the two-parameter logistic model. The x_{n+1} and x_n obtained in this manner cannot be related in an exact manner like (15). By using linear approximations, it was shown (Wu, 1985) that $x_{n+1} - x_n$ can be approximated by the adaptive RM scheme (1) with $c = \hat{\beta}_n^{-1}$. On the other hand, a one-parameter model does not lead to a recursive scheme asymptotically equivalent to the adaptive RM scheme. This point should be clear from the discussion in Section 2, (3) and (4). For binary data, the recursion (15) based on the one-parameter logistic model again can not be approximated by the adaptive RM. This is because the slope parameter in the two-parameter model plays the role of the regression slope parameter β in the adaptive RM. Without a consistent estimate of the slope parameter, first-order optimality of the ML recursion in terms of minimizing a_1 in (13) cannot be achieved.

The logistic model has a unique place for binomial data in that its likelihood equation resembles the normal equation in linear models. For generalized linear models to be discussed in Section 4, this unique role is played by the canonical link function.

Since the RM recursion has desirable asymptotic properties under weak conditions on M and the ML recursion can be approximated by a version of

the RM scheme, this nonparametric property may be shared by the ML recursion. In the following we will use the logistic model to explain why the ML recursion approach, apparently model-based, is robust against model misspecification. The following argument is taken from Wu (1985). The likelihood equation for α is

$$(16) \qquad n^{-1} \sum_{1}^{n} y_i = n^{-1} \sum_{1}^{n} H(x_i|\hat{\alpha}_n,\hat{\lambda}_n), H(x|\alpha,\lambda) = (1 + e^{-\lambda x+\alpha})^{-1} .$$

Make a rather strong assumption that $\hat{\alpha}_n \to \alpha^*$, $\hat{\lambda}_n \to \lambda^*$ uniformly so that x_{n+1}, which satisfies $H(x_{n+1}|\hat{\alpha}_n,\hat{\lambda}_n) = p$, converges to a constant w. The ML recursion, based on the H function, is robust if w satisfies $M(w) = p$. Recall that M is the unknown true response function which may be different from H. For continuous M, the left side of (16) converges to $M(w)$ a.s., since each y_i is binomial with probability $M(x_i)$. This side does <u>not</u> depend on the H assumption. The right side of (16) converges to $H(w|\alpha^*,\lambda^*) = \lim H(x_{n+1}|\hat{\alpha}_n,\hat{\lambda}_n) = p$. Therefore $M(w) = p$. The right side, though starting with the H assumption, turns out to be equal to the constant p because of the recursion step

$$(17) \qquad H(x_{n+1}|\hat{\alpha}_n,\hat{\lambda}_n) = p ,$$

which is recognized as the <u>source of robustness</u> of the ML recursion approach. In other words, a possible misspecification in H is "undone" by the recursion step (17). This robustness claim may not hold for the estimation of other parameters such as the slope of M at x^*.

Another interpretation is that the assumed model H is locally valid in x, whatever the true "global" model is.

4. EXTENSIONS TO GENERALIZED LINEAR MODELS

The ML recursion approach to sequential design can be applied to very general variations described by a generalized linear model. The response y has a density function

$$\exp\{[y\theta - b(\theta)]/a(\varphi) + c(y,\varphi)\}$$

for some functions a,b,c. If φ is known, this is an exponential family with canonical parameter θ. The response y is related to the variable x through a link function $\eta : \theta \to \eta(\theta)$ such that the η scale is linear in x, i.e., $\eta = \lambda x - \alpha$. Typically the link function is unknown. The mean response $M(x) = E(y|x)$ is

$$M(x) = b'(\theta) = b'(\eta^{-1}(\lambda x - \alpha)) .$$

Without knowing M, we assume a link function $\xi:\theta \to \xi(\theta)$ so that $\xi = \lambda x - \alpha$ and the mean response function is

$$H(x|\alpha,\lambda) = b'(\theta) = b'(\xi^{-1}(\lambda x - \alpha)) ,$$

where ξ^{-1} is the inverse function of ξ. The likelihood equation obtained by differentiating, with respect to α and λ,

$$\sum y_i \xi^{-1}(\lambda x_i - \alpha) = \sum b(\xi^{-1}(\lambda x_i - \alpha)) ,$$

is, by writing $\xi^{-1} = \rho$,

$$\begin{aligned} \sum y_i \rho'(\lambda x_i - \alpha) &= \sum H(x_i|\alpha,\lambda)\rho'(\lambda x_i - \alpha) \\ \sum x_i y_i \rho'(\lambda x_i - \alpha) &= \sum x_i H(x_i|\alpha,\lambda)\rho'(\lambda x_i - \alpha) . \end{aligned} \tag{18}$$

If $\xi(\theta) = \theta$, $\sum y_i$ and $\sum y_i x_i$ are the sufficient statistics for α and λ and (18) resembles the normal equation in linear models. Such a link function is called a canonical link (McCullagh and Nelder, 1983). For $N(\mu,\sigma^2)$, $\xi(u) = \mu$ is the canonical link; for binomial with probability p, the logit function $\xi(p) = \log[p(1 - p)^{-1}]$ is the canonical link and $\xi(p) = \lambda x - \alpha$ gives the logistic function $p = (1 + e^{-\lambda x+\alpha})^{-1}$. Without

a priori reason for choosing other H functions, the canonical link function is a convenient choice for the ML recursion approach.

For estimating x^* with $M(x^*) = p$, where p is in the range of the mean $b'(\theta)$, the ML recursion works as follows. Let $\hat{\alpha}_n$ and $\hat{\lambda}_n$ be a solution to (18) based on the first n observations. Take the next observation y_{n+1} at x_{n+1} satisfying

$$H(x_{n+1}|\hat{\alpha}_n,\hat{\lambda}_n) = b'(\xi^{-1}(\hat{\lambda}_n x_{n+1} - \hat{\alpha}_n)) = p \ .$$

In the next section, we shall study another special case, the Poisson distribution.

5. SEQUENTIAL POISSON EXPERIMENTS

Examples of Poisson variation include radiation counts and number of jobs arriving in a period. The associated x variable may be the distance to the source of radiation or the parameter specification of a queuing system. Here y is a Poisson variable with mean μ. To estimate x^* with $M(x^*) = E(y|x^*) = p$, we assume a canonical link with one parameter, i.e.,

$$p(y|\mu) \propto \exp(y \ln \mu - \mu), \quad \ln \mu = x - \alpha \ . \tag{19}$$

The mean response function according to (19) is $H(x|\alpha) = e^{x-\alpha}$. The parameter of interest is $\tilde{x}$ satisfying $H(\tilde{x}|\alpha) = p$, i.e. $e^{\tilde{x}} = pe^{\alpha}$. By solving the equation

$$\frac{\partial}{\partial \alpha} \sum_{1}^{n} [y_i(x_i - \alpha) - e^{x_i - \alpha}] = 0 \ ,$$

we obtain the maximum likelihood estimate $\hat{\alpha}_n$ of α, through

$$e^{-\hat{\alpha}_n} = \sum^{n} y_i \ / \sum^{n} e^{x_i} \ .$$

If the next design x_{n+1} is chosen to satisfy $H(x_{n+1}|\hat{\alpha}_n) = p$,

$$e^{x_{n+1}} = pe^{\hat{\alpha}_n} = p \sum^{n} e^{x_i} \ / \sum^{n} y_i \ . \tag{20}$$

Equation (20) for n and $n-1$ gives

$$e^{x_{n+1}} - e^{x_n} = p \frac{e^{x_n} \Sigma^{n-1} y_i - y_n \Sigma^{n-1} e^{x_i}}{\Sigma^{n} y_i \Sigma^{n-1} y_i} = p \frac{p - y_n}{e^{-\hat{\alpha}_{n-1}} \Sigma^{n} y_i} \ ,$$

or equivalently,

$$e^{x_{n+1}} = e^{x_n} - \frac{c_n}{n}(y_n - p), \qquad c_n^{-1} = \left(\frac{1}{n} \sum^{n} y_i\right)\left(p^{-1} e^{-\hat{\alpha}_{n-1}}\right) \ . \tag{21}$$

Equation (21) is an RM recursion in the <u>transformed</u> scale e^x.

By rewriting the ML recursion as an RM recursion, we can draw on the rich literature on the asymptotic results of the latter procedure. The remaining

section is devoted to a heuristic study of the limiting behavior of (21). Assume that x_n converges to a constant z_0. From the martingale strong law of large numbers,

$$\frac{1}{n}\sum^{n} y_i \approx \frac{1}{n}\sum^{n} M(x_i) \to M(z_0) \quad . \tag{22}$$

Similarly, by letting $n \to \infty$ on both sides of (20),

$$e^{z_0} = p\,\frac{e^{z_0}}{M(z_0)} \quad ,$$

implying that $M(z_0) = p$ and $z_0 = x^*$, that is x_n is asymptotically consistent. The asymptotic variance of e^{x_n} to e^{x^*} depends on the limiting value of c_n in (21). From (20) and (22),

$$c_n^{-1} + M(z_0)e^{-z_0} = pe^{-x^*} \quad ,$$

which is not equal to $M'(x^*)$ and therefore the scheme (21) does not have minimal asymptotic variance. A rigorous treatment of the consistency and asymptotic normality of e^{x_n} in (21) is desired. A general result of Sellke (1985) may be applicable.

6. CONCLUDING REMARKS

The ML recursion approach to sequential designs is based on ideas well known to statisticians. It is intuitively appealing and easy to understand, although the stochastic approximation is slightly easier to implement. It is applicable to very general distributions. Special features such as discreteness or boundedness of the data are taken into account through a proper choice of the likelihood. In several important situations, it is very close to the RM recursive scheme (with varying choice of the constant c) and thus shares the latter's robustness against misspecified link function. If the assumed model is correct, it is asymptoticaly efficient and may also perform well in small samples as the maximum likelihood estimator often does. Wu's (1985) simulation results for binary data suggest that it may be superior to the RM recursion in small samples.

Its major problem thus far is the lack of rigorous theory on its asymptotic behavior. An attempt has been made by Sellke (1985). Here its linkage to the RM scheme pays off. By rewriting it as an RM-like scheme and suitably bounding the constant c in (1), simple proofs of its asymptotic properties may be obtained by drawing on the vast literature on the RM scheme. Another gain due to this linkage is in the choice of stabilizing constant. Adaptive versions of the ML or RM recursion, though asymptotically optimal, may not perform well in small samples because of the instability caused by adaptation. Stability can be achieved by putting a bound on the constant c in (1). The simulation results of Wu (1985) demonstrate the effectiveness of this device in reducing the small-sample mean square errors of both procedures. Cross-fertilization of the two approaches may lead to further understanding and results.

Acknowledgement. I would like to thank T. Sellke, D. M. Titterington and E. W. Frees for helpful comments.

REFERENCES

Anbar, D. (1978). A stochastic Newton-Raphson method. J. Statist. Planning and Inference 2, 153-163.

Goodwin, G. C. and Payne, R. L. [1977). Dynamic System Identification: Experimental Design and Data Analysis. Academic Press, New York.

Lai, T. L. and Robbins, H. (1979). Adaptive design and stochastic approximation. Ann. Statist. 7, 1196-1221.

Lai, T. L. and Robbins, H. (1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation scheme. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete. 56, 329-360.

Lai, T. L. and Robbins, H. (1982). Iterated least squares in multiperiod control. Advances in Applied Mathematics 3, 50-73.

McCullagh, P. and Nelder, J. A. (1983). Generalized Linear Models. Chapman and Hall, London.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. Ann. Math. Statist. 29, 400-407.

Robbins, H. and Siegmund, D. (1971). A convergence theorem for non-negative almost sure supermartingale and some applications. Optimization Methods in Statistics (ed. J. N. Rustagi), Academic Press, 233-257.

Sacks, J. (1958). Asymptotic distribution of stochastic approximation procedures. Ann. Math. Statist. 29, 373-405.

Sellke, T. (1985). Consistency of maximum likelihood recursion in stochastic approximation. Purdue University Technical Report, to appear.

Sellke, T. (1986). Repeated-MLE procedures for stochastic approximation in quantal response problems. This volume.

Wu, C. F. J. (1985). Efficient sequential designs with binary data. J. Amer. Statist. Assoc. 80, December, to appear.

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2890	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* MAXIMUM LIKELIHOOD RECURSION AND STOCHASTIC APPROXIMATION IN SEQUENTIAL DESIGNS		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR*(s)* C. F. J. Wu		8. CONTRACT OR GRANT NUMBER*(s)* DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53705		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE November 1985 13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

maximum likelihood recursion
stochastic approximation
Robbins-Monro
binary experiment
Poisson experiment
generalized linear model
canonical link

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

An approach to sequential design for estimating the root of a nonlinear equation is described. It sets the next design point at the current estimate of the parameter via a parametric model and maximum likelihood (or other efficient) estimation. For normal, binomial and Poisson errors and their respective canonical link functions, it is close to the Robbins-Monro stochastic approximation and thus enjoys the latter's robustness against the misspecification of the link function. Some new variations of the Robbins-Monro scheme are obtained as a consequence.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

END

FILMED

3 -86

DTIC